

Linear Transmission of Composite Gaussian Measurements over a Fading Channel under Delay Constraints

Onur Tan*, Deniz Gündüz[†], Jesús Gómez Vilardebò*

*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain.

[†]Imperial College London, London, UK.

Emails: {onur.tan@cttc.es, d.gunduz@imperial.ac.uk, jesus.gomez@cttc.es}

Abstract

Delay constrained linear transmission (LT) strategies are considered for the transmission of composite Gaussian measurements over an additive white Gaussian noise fading channel under an average power constraint. If the channel state information (CSI) is known by both the encoder and decoder, the optimal LT scheme in terms of the average mean-square error distortion is characterized under a strict delay constraint, and a graphical interpretation of the optimal power allocation strategy is presented. Then, for general delay constraints, two LT strategies are proposed based on the solution to a particular multiple measurements-parallel channels scenario. It is shown that the distortion decreases as the delay constraint is relaxed, and when the delay constraint is completely removed, both strategies achieve the optimal performance under certain matching conditions. If the CSI is known only by the decoder, the optimal LT strategy is derived under a strict delay constraint. The extension for general delay constraints is shown to be hard. As a first step towards understanding the structure of the optimal scheme in this case, it is shown that for the multiple measurements-parallel channels scenario, any LT scheme that uses only a one-to-one linear mapping between measurements and channels is suboptimal in general.

Index Terms

Linear transmission, delay constraint, composite of Gaussians, fading channel, water filling, wireless sensor networks.

This work was supported by the Spanish Government under project TEC2013-44591-P (INTENSIV).

I. INTRODUCTION

Near real-time monitoring of a physical phenomena is of great significance to many emerging network applications, such as monitoring of voltage, current magnitudes, active/reactive power values in smart grids (SGs) [1], or temperature and humidity in forest fire detection networks [2]. To this end, wireless sensors are deployed throughout the physical network and the sensor measurements are delivered to a control center (CC) over wireless links. For the robust, reliable and efficient management of the underlying physical networks, near real-time and accurate reconstruction of the measurements at the CC becomes necessary. For example, in conventional state estimation for the electricity grid, measurements are collected once every two to four seconds and the state is updated once every few minutes [3]. However, more frequent state measurements and estimations are required for modern SGs, which inevitably imposes strict delay constraints on the transmission of measurements. As a further example, in forest fire detection networks [4], measurements of smoke and gas sensors along with camera images are used to detect fire, and the delay inevitably becomes a major constraint for the transmission. Thus, zero-delay linear transmission (LT), rather than advanced compression and channel coding techniques that span large codewords, is an attractive strategy for the transmission of sensor measurements in intelligent networks. This is because LT reduces both the delay and encoding complexity significantly; and accordingly limits the cost and energy requirements of the sensors.

LT of Gaussian sources has been extensively studied in the literature. Goblick showed in [5] that zero-delay LT of a Gaussian source over an additive white Gaussian noise (AWGN) channel achieves the optimal mean-square error (MSE) distortion. In [6], the optimal LT scheme that matches an r -dimensional Gaussian signal to a k -dimensional AWGN vector channel is characterized. It is shown that the optimal LT performance can be achieved by mapping ordered sources to ordered channels in a one-to-one fashion. LT of a Gaussian source over a fading AWGN channel is studied in [7]. It is shown that the optimal LT performance can be achieved by symbol-by-symbol processing, and increasing the block length does not provide any gain, as opposed to nonlinear coding schemes. In [8], LT of noisy vector measurements over a fading AWGN channel is studied under diagonal and general observation matrices. LT of vector Gaussian sources over static and fading multi-antenna channels is studied in [9] and [10], respectively.

We consider a wireless sensor node that collects measurements from J Gaussian parameters.

We discretize time into time slots (TSs), and assume that the CC asks for a measurement of a particular parameter from the sensor at each TS. The sensor takes one sample of the requested parameter at each TS, and transmits these samples to the CC over an AWGN fading channel under a given delay constraint. Note that, in contrast to multi-dimensional Gaussian source models studied in [6], [10], [11], where the sensor has the measurements of all the J Gaussian parameters at the beginning of a TS, we assume that only one measurement is taken from the requested parameter at each TS.

We assume that each measurement must be delivered within d TSs. Thereby, after each transmission, the CC estimates the measurement whose deadline is just about to expire. We assume that the channel gain from the sensor to the CC is independent and identically distributed (i.i.d.) over TSs. We consider two different scenarios regarding the channel state information (CSI): In the first scenario, the CSI is assumed to be available to both the encoder and decoder, while in the second scenario, only the decoder has CSI. Our goal is to estimate all the requested measurements at the CC within their delay constraints with the minimum MSE distortion.

We focus explicitly on LT strategies. Assuming that the CSI is known by both the encoder and decoder, we first derive the optimal LT strategy under a strict delay constraint ($d = 1$), and show that the optimal power allocation and the corresponding distortion can be interpreted as *water-filling reflected on a reciprocal mirror*. Exploiting the results of [6], we also derive the optimal LT strategy under a strict delay constraint for a particular scenario in which the sensor transmits the measurement vector over parallel AWGN fading channels at each TS. Then, building on our previous works [12], [13], and exploiting the optimal LT strategy derived for multiple measurements-parallel channels scenario above, we propose two LT strategies for general delay constraints. In both strategies, measurements are first collected and stored in a buffer whose size depends on the delay constraint, and then, are transmitted to the CC over multiple channel accesses within the delay constraint. The two strategies consider different measurement selection criterias, which are used to select the appropriate stored measurement to be transmitted at each channel access. We then derive the theoretical lower bound (TLB) and the LT lower bound (LLB) on the achievable MSE distortion. We characterize the MSE distortion achieved by the proposed LT schemes, as well as the TLB and the LLB under various power and delay constraints. We show that the MSE distortion diminishes as the delay constraint is relaxed if the sensor is capable of measuring more than one system parameter, i.e., $J > 1$. However, if $J = 1$, then relaxing the

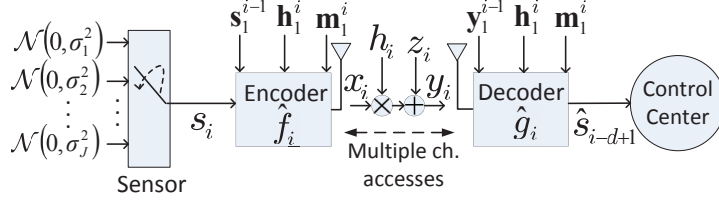


Fig. 1. The illustration of the transmission model from the perspective of the sensor with multiple channel accesses.

delay constraint does not provide any improvement in LT performance as argued in [6]. When the fading channel follows a discrete distribution and the delay constraint is completely removed, we show that the proposed LT strategies meet the TLB under certain matching conditions between the channel states and the parameter variances; and hence, achieve the optimal performance.

When the CSI is known only by the decoder, we first derive the optimal LT strategy under a strict delay constraint. Then, we consider the multiple measurements-parallel channels scenario under a strict delay constraint and $J > 1$ assumption, and show that the optimal LT performance cannot be achieved by an LT scheme that is constrained to use only a one-to-one linear mapping between measurements and channels, as opposed to the $J = 1$ case [7], and the CSI is known by both the encoder and decoder [6], respectively. Since the optimal LT strategy is elusive for $J > 1$, we do not consider LT strategies for larger delay constraints. Finally, we derive the TLB on the achievable MSE distortion.

The rest of the paper is structured as follows. The system model is presented in Section II. In Sections III to V CSI is assumed at both the encoder and decoder. In Section III, we study the optimal LT strategy under a strict delay constraint. Two LT strategies are proposed for general delay constraints in Section IV. In Section V, we characterize the TLB and LLB on the achievable MSE distortion. In Section VI, the optimal LT strategy is derived under a strict delay constraint along with the TLB, when the CSI is known only by the decoder. Section VII presents extensive numerical results, and finally, Section VIII concludes the paper.

II. SYSTEM MODEL

We consider a CC that monitors the operation of a system through a wireless sensor (Fig. 1), which is capable of measuring J distinct system parameters. The j th system parameter is modelled as a zero-mean Gaussian random variable (r.v.) with variance σ_j^2 , i.e., $\mathcal{N}(0, \sigma_j^2)$, for

$j \in [1:J]$, where $[1:J]$ denotes the set $\{1, 2, \dots, J\}$. These system parameters are independent from each other, and their realizations are i.i.d. over time. In order to monitor the network operation, the CC requests the measurement of one system parameter from the sensor at each TS. The index of the requested system parameter at each TS is a r.v. denoted by $M \in [1:J]$, with distribution $p_M(m)$, which is also i.i.d. over time. Based on these requests, the sensor takes one measurement of the requested parameter m at each TS. Thereby, the model is that of a composite source introduced in Chapter 6 of [14]. The source S can be described as a composite source comprised of J distinct components (subsources), each operating independently of the others. In our model, each component produces data according to a Gaussian probability distribution $P(\cdot|m) = \mathcal{N}(0, \sigma_m^2)$. The set G of all sources comprises the composite source $\{S_i, G\}$. In our case,

$$G = [\mathcal{N}(0, \sigma_1^2), \mathcal{N}(0, \sigma_2^2), \dots, \mathcal{N}(0, \sigma_J^2)] . \quad (1)$$

The index of the requested system parameter m generates the sequence of positions assumed by the switch in Fig. 1. In our model both the encoder and the decoder possess the exact knowledge of this sequence. Notice that, in the particular case in which the encoder and decoder are uninformed about this sequence, the composite source $\{S_i, G\}$ is equivalent to a mixture of Gaussian distributions, i.e., $P_S(s) = \sum_{m=1}^J P_M(m) P_{S|M}(s|m)$.

We assume that the CC imposes a maximum delay constraint of $d \in \mathbb{Z}^+$ on the measurements, that is, the measurement requested in a TS needs to be transmitted within the following d TSs; otherwise, it becomes stale. The collected sensor measurements are transmitted to the CC over a fading channel with zero-mean and unit variance AWGN. The channel output at TS i is given by $y_i = h_i x_i + z_i$, where x_i is the channel input, z_i is the additive noise with $Z \sim \mathcal{N}(0, 1)$, and h_i is the fading state of the channel. We consider a fading channel model, and assume that the fading coefficient $H_i \in \mathbb{R}$ is modelled as a r.v. i.i.d. over time with probability distribution $p_H(h)$.

We define $\mathbf{m}_i^l = [m_i, m_{i+1}, \dots, m_l]$ as the sequence of indices of requested parameters at TSs $[i:l]$ for $i \leq l$. The measurement sequence is defined similarly as $\mathbf{s}_i^l = [s_i, \dots, s_l]$, where the i -th entry s_i is the measured value of the requested parameter m_i at TS i . Therefore, the sequence \mathbf{s}_i^l has independent entries, where the i -th entry comes from a Gaussian distribution with variance $\sigma_{m_i}^2$. Note that in our composite Gaussian measurements model, conditioned on the requested

parameter index, which is known by both the encoder and decoder, the source samples follow Gaussian distributions with different variance values.

The channel state and the output sequences at TSs $[i:l]$ are similarly defined as $\mathbf{h}_i^l = [h_i, \dots, h_l]$ and $\mathbf{y}_i^l = [y_i, \dots, y_l]$, respectively. We assume that both the encoder and decoder at TS i know all the past channel states, \mathbf{h}_1^{i-1} , and the indices of requested parameters, \mathbf{m}_1^i , as well as the statistics of the measured parameters, σ_m^2 , the parameter requests, $p_M(m)$, and the channel, $p_H(h)$. In the first part of the paper we assume that both the encoder and decoder know the current channel state, h_i . Note that this assumption might be hard to realize for a fast fading channel model; on the other hand, our system model can be considered as instances of a slow fading channel. Typically, there will be a large number of sensors in the system, and each sensor is going to be scheduled only once in a while; and hence, each TS in our system model can be considered as one instance of a slow fading channel when a particular sensor is scheduled to transmit. Since these instances are separated from each other due to the transmission of other sensors, corresponding channel states are modeled as i.i.d., and are assumed to be known by both the encoder and decoder, as channel estimation and CSI feedback can be carried out between two transmissions of the same sensor. In Section VI we will consider the scenario in which the CSI is known only by the decoder.

1) *Encoding Function:* The encoding function $\hat{f}_i : \mathbb{R}^i \times \mathbb{R}^i \times \mathbb{R}^i \rightarrow \mathbb{R}$, maps \mathbf{s}_1^i , \mathbf{h}_1^i , and \mathbf{m}_1^i to a channel input x_i at each TS i , i.e., $x_i = \hat{f}_i(\mathbf{s}_1^i, \mathbf{h}_1^i, \mathbf{m}_1^i)$. An average power constraint of P is imposed on the encoding function:

$$\bar{P} \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{M,H,S} [|X_i|^2] \leq P,$$

where $\mathbb{E}_{M,H,S} [\cdot]$ denotes the expectation over M , H and S . For any generic transmission policy, the encoding function \hat{f}_i , at TS i , may consider to use any combination of \mathbf{s}_1^i , \mathbf{h}_1^i , and \mathbf{m}_1^i to generate x_i . This gives rise to a time-varying encoding scheme.

2) *Decoding Function:* At the end of TS i , the goal of the CC is to estimate with the minimum MSE distortion, the measurement s_{i-d+1} , which has been requested exactly $d-1$ TSs ago, and is about to expire. The decoding function $\hat{g}_i : \mathbb{R}^i \times \mathbb{R}^i \times \mathbb{R}^i \rightarrow \mathbb{R}$, estimates \hat{s}_{i-d+1} based on \mathbf{y}_1^i , \mathbf{h}_1^i , and \mathbf{m}_1^i , i.e., $\hat{s}_{i-d+1} = \hat{g}_i(\mathbf{y}_1^i, \mathbf{h}_1^i, \mathbf{m}_1^i)$. The MSE distortion is given by:

$$\bar{D} \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=d}^n \mathbb{E}_{M,H,S,Z} \left[|S_{i-d+1} - \hat{S}_{i-d+1}|^2 \right].$$

The decoding function \hat{g}_i , at TS i , reconstructs the measurement using \mathbf{y}_1^i , \mathbf{h}_1^i , and \mathbf{m}_1^i . Hence, similarly to the encoder, the decoder may be time-varying.

We are interested only in LT policies in which \hat{f}_i 's are restricted to be linear functions of the sensor measurements, s_i 's, i.e., $\hat{f}_i(\mathbf{s}_1^i, \mathbf{h}_1^i, \mathbf{m}_1^i) \triangleq f_i(\mathbf{h}_1^i, \mathbf{m}_1^i) \cdot \mathbf{s}_1^i$, where $f_i(\mathbf{h}_1^i, \mathbf{m}_1^i)$ is a vector. Under this linearity constraint, the optimal estimators at the receiver, \hat{g}_i 's, are also linear functions of the channel outputs, y_i 's, i.e., $\hat{g}_i(\mathbf{y}_1^i, \mathbf{h}_1^i, \mathbf{m}_1^i) \triangleq g_i(\mathbf{h}_1^i, \mathbf{m}_1^i) \cdot \mathbf{y}_1^i$, where $g_i(\mathbf{h}_1^i, \mathbf{m}_1^i)$ is a vector. Hereafter, we will refer to f_i and g_i for the encoding and decoding functions at TS i , respectively.

III. STRICT DELAY CONSTRAINT

We first derive the optimal LT strategy under a strict delay constraint ($d = 1$), and characterize the minimum achievable MSE distortion. In this scenario, optimal LT performance is achieved by transmitting only the most recent measurement since all the previous measurements have expired, and transmitting an expired measurement cannot help the estimation of the current measurement since the measurements are independent. Then the encoding function $f_i(h_i, m_i)$ at TS i is a scalar. Given the encoding function, the decoding function $g_i(h_i, m_i)$ that minimizes the MSE for Gaussian r.v.s is the linear MMSE estimator [15], and is also a scalar.

In particular, for a measurement s_i with variance $\sigma_{m_i}^2$, and channel output $y_i = h_i \cdot f_i(h_i, m_i) \cdot s_i + z_i$ at TS i , the decoding function can be written explicitly as:

$$g_i(h_i, m_i) = \frac{\mathbb{E}_{S,Z}[S_i Y_i]}{\mathbb{E}_{S,Z}[Y_i^2]} = \frac{|h_i| f_i(h_i, m_i) \sigma_{m_i}^2}{|h_i|^2 f_i(h_i, m_i)^2 \sigma_{m_i}^2 + 1}. \quad (2)$$

In the following lemma we show that there is no loss of optimality by limiting the encoding function to be time-invariant.

Lemma 1. *Under a strict delay constraint there is no loss of optimality by considering only time-invariant encoding functions, i.e., $f_i(h_i, m_i) = f(h_i, m_i) \forall i$.*

Proof:

$$\bar{D} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{M,H,S,Z} [|S_i - \hat{S}_i|^2] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{M,H} \left[\frac{\sigma_m^2}{|h|^2 f_i(h, m)^2 \sigma_m^2 + 1} \right], \quad (3)$$

$$\geq \mathbb{E}_{M,H} \left[\frac{\sigma_m^2}{|h|^2 f(h, m)^2 \sigma_m^2 + 1} \right], \quad (4)$$

where (3) is the average MSE distortion under a strict delay constraint ($d = 1$); and defining $f(h, m)^2 \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_i(h, m)^2$ such that $f(h, m)$ satisfies the average power constraint P , (4) follows from the convexity of the function $\mathbb{E}_{M,H} \left[\frac{\sigma_m^2}{|h|^2 f_i(h, m)^2 \sigma_m^2 + 1} \right]$ in terms of $f_i(h, m)^2$, and the equality holds iff $f_i(h, m) = f(h, m)$ for $\forall i$ and due to the strict convexity of the aforementioned function. Thus, the time-invariant encoding function $f(h, m)$, which is a function of only h and σ_m^2 , does not lead to any loss in optimality. \blacksquare

The time-invariant encoding function $f(h, m)$ leads to a time-invariant decoding function $g(h, m)$. In the rest of the paper, we will consider time-invariant encoding and decoding functions without loss of optimality. Then, the MSE distortion, $\bar{D} = \mathbb{E}_{M,H,S,Z} [|S - \hat{S}|^2]$, and the average power, $\bar{P} = \mathbb{E}_{M,H,S} [|X|^2]$, can be written explicitly as functions of h and σ_m^2 , as follows:

$$\bar{D} = \sum_{m=1}^J p_M(m) \int_{\mathbb{R}} \frac{\sigma_m^2}{|h|^2 f(h, m)^2 \sigma_m^2 + 1} p_H(h) dh, \quad (5)$$

$$\bar{P} = \sum_{m=1}^J p_M(m) \int_{\mathbb{R}} f(h, m)^2 \sigma_m^2 p_H(h) dh. \quad (6)$$

The optimal linear encoding function $f^*(h, m)$ is found as the solution to the convex optimization problem $\bar{D}^* \triangleq \min_f \bar{D}$, subject to the average power constraint $\bar{P} \leq P$. From the Karush-Kuhn-Tucker (KKT) optimality conditions [16], we obtain:

$$f^*(h, m) = \sqrt{\left[\frac{\lambda^*}{|h| \sigma_m} - \frac{1}{|h|^2 \sigma_m^2} \right]^+}, \quad (7)$$

where λ^* is the optimal Lagrange multiplier, such that $\bar{P} = P$.

The optimal power allocation and the corresponding distortion are given by:

$$P^*(h, m) = \frac{\sigma_m}{|h|} \left[\lambda^* - \frac{1}{|h| \sigma_m} \right]^+, \quad (8)$$

$$D^*(h, m) = \frac{\sigma_m}{|h|} \min \left(\frac{1}{\lambda^*}, |h| \sigma_m \right), \quad (9)$$

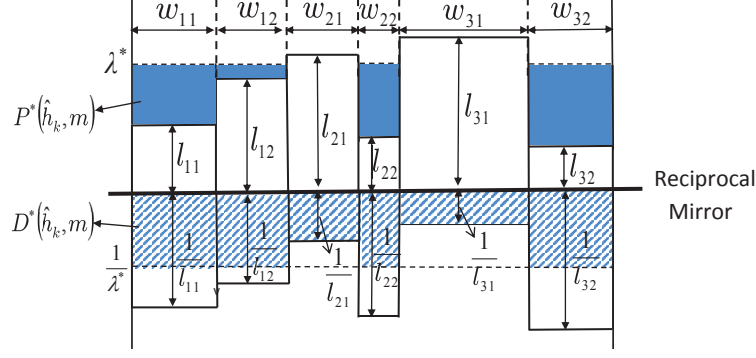


Fig. 2. Water-filling reflected on a reciprocal mirror.

where $\bar{D}^* = \mathbb{E}_{M,H} [D^*(h, m)]$ and $\mathbb{E}_{M,H} [P^*(h, m)] = P$.

In Fig. 2, we present a graphical interpretation of the optimal power allocation and the corresponding distortion for $J = 2$ parameters with variances σ_1^2 and σ_2^2 , which are requested with probabilities $p_M(1), p_M(2)$, respectively. We also consider a discrete fading channel with three states, where the k th state, \hat{h}_k , is observed with probability $p_H(\hat{h}_k)$, $k = 1, 2, 3$. Fig. 2 depicts rectangles that are placed upon a mirror surface and their reciprocally scaled images below. Rectangles represent all possible source-channel pairs $\{\sigma_m, \hat{h}_k\}$, where $l_{km} \triangleq \frac{1}{|\hat{h}_k| \sigma_m}$ and $w_{km} \triangleq \frac{\sigma_m}{|\hat{h}_k|}$ indicate the height and width of the rectangles, respectively. The total power is poured above the level l_{km} up to the water level λ^* across the rectangles placed upon the mirror. The optimal power allocated to the source-channel pair $\{\sigma_m, \hat{h}_k\}$ is given by the shaded area below the water level and above l_{km} . The corresponding distortion values are found by simply looking at the reciprocally scaled reflections of the rectangles and the water level on the mirror. If $\frac{1}{l_{km}} > \frac{1}{\lambda^*}$, distortion is given by the width w_{km} times the reciprocal of the water level $\frac{1}{\lambda^*}$, and if $\frac{1}{l_{km}} \leq \frac{1}{\lambda^*}$, distortion is σ_m^2 , which are illustrated as dashed areas in Fig. 2. We call this as *water-filling reflected on a reciprocal mirror*.

A. Multiple Measurements and Parallel Channels

Next, we assume that the CC requests $N > 1$ measurements from the sensor at each TS, and the sensor transmits a length- N measurement vector over N parallel orthogonal AWGN fading channels under a strict delay constraint ($d = 1$). For this scenario, we characterize the optimal LT strategy by generalizing the results of [6] derived for Gaussian vector sources to our composite Gaussian measurements model. This scenario differs from the system model defined

in Section II, since we allow to take N measurements at each TS as opposed to taking only one measurement at each TS. However, we will exploit the optimal LT strategy in this setting for the construction of the proposed transmission strategies in Section IV, as well as for characterizing the LLB in Section V-B.

Only for this scenario, we define $\mathbf{m} = [m_1, \dots, m_N]$ as the vector of indices of N requested parameters at a particular TS. Then, the sensor takes the length- N measurement vector $\mathbf{s} = [s_1, \dots, s_N]$ according to the parameters indicated by \mathbf{m} , i.e., s_1 is the measured value of parameter m_1 . For a strict delay constraint ($d = 1$), the optimal LT performance is achieved by transmitting only the most recent measurement vector. Similarly to Lemma 1, the encoding function can be limited to a time-invariant $N \times N$ square matrix $\mathbf{F}_{\mathbf{h}, \mathbf{m}}$ without loss of optimality, where subscripts \mathbf{h} and \mathbf{m} indicate the dependence of the encoding matrix on the realizations of \mathbf{h} and \mathbf{m} . We assume that the N channels are i.i.d with distribution $p_H(h)$, and \mathbf{H} is defined as the $N \times N$ diagonal channel matrix. The diagonal elements of \mathbf{H} are denoted by a channel vector $\mathbf{h} = [h_1, \dots, h_N]$ at a particular TS. The length- N channel output vector at that particular TS is given by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$, where \mathbf{x} is the length- N channel input vector and \mathbf{z} is the length- N additive noise vector with $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

The encoder at any TS maps its measurement vector \mathbf{s} , to a channel input vector \mathbf{x} , i.e., $\mathbf{x} = \mathbf{F}_{\mathbf{h}, \mathbf{m}}\mathbf{s}$. An average power constraint of P is imposed on the encoding function:

$$\bar{P} = \frac{1}{N} \text{Tr} \{ \mathbb{E}_{M,H,S} [\mathbf{x}\mathbf{x}^T] \} = \frac{1}{N} \text{Tr} \{ \mathbb{E}_{M,H} [\mathbf{F}_{\mathbf{h}, \mathbf{m}} \mathbf{C}_s \mathbf{F}_{\mathbf{h}, \mathbf{m}}^T] \} \leq P, \quad (10)$$

where $\mathbf{C}_s = \mathbb{E}_S [\mathbf{s}\mathbf{s}^T]$.

Given the encoding function, the decoding function that minimizes the MSE for a Gaussian random vector is the $N \times N$ linear MMSE estimator matrix $\mathbf{G}_{\mathbf{h}, \mathbf{m}}$ [15], which is also time-invariant. Similarly to $\mathbf{F}_{\mathbf{h}, \mathbf{m}}$, subscripts \mathbf{h} and \mathbf{m} indicate the dependence of the decoding matrix on the realizations of \mathbf{h} and \mathbf{m} . For the measurement vector \mathbf{s} , and the channel output vector \mathbf{y} , at any TS, we have:

$$\mathbf{G}_{\mathbf{h}, \mathbf{m}} = \mathbf{C}_{\mathbf{s}\mathbf{y}} \mathbf{C}_{\mathbf{y}}^{-1} = \mathbf{C}_s \mathbf{F}_{\mathbf{h}, \mathbf{m}}^T \mathbf{H}^T \Phi, \quad (11)$$

where $\mathbf{C}_{\mathbf{s}\mathbf{y}} = \mathbb{E}_{S,Z} [\mathbf{s}\mathbf{y}^T]$, $\mathbf{C}_{\mathbf{y}} = \mathbb{E}_{S,Z} [\mathbf{y}\mathbf{y}^T]$ and $\Phi \triangleq (\mathbf{H}\mathbf{F}_{\mathbf{h}, \mathbf{m}} \mathbf{C}_s \mathbf{F}_{\mathbf{h}, \mathbf{m}}^T \mathbf{H}^T + \mathbf{I})^{-1}$.

At any TS, the CC estimates the most recent measurement vector \mathbf{s} as $\hat{\mathbf{s}}$, i.e., $\hat{\mathbf{s}} = \mathbf{G}_{\mathbf{h}, \mathbf{m}}\mathbf{y}$. The MSE distortion is given by:

$$\bar{D} = \frac{1}{N} \text{Tr} \{ \mathbb{E}_{M,H,S,Z} [|\mathbf{s} - \hat{\mathbf{s}}| |\mathbf{s} - \hat{\mathbf{s}}|^T] \} = \frac{1}{N} \text{Tr} \{ \mathbb{E}_{M,H} [\mathbf{C}_s - \mathbf{C}_s \mathbf{F}_{\mathbf{h},\mathbf{m}}^T \mathbf{H}^T \Phi \mathbf{H} \mathbf{F}_{\mathbf{h},\mathbf{m}} \mathbf{C}_s] \}. \quad (12)$$

The optimal linear encoding matrix $\mathbf{F}_{\mathbf{h},\mathbf{m}}^*$, is found as the solution to the convex optimization problem $\bar{D}^* \triangleq \min_{\mathbf{F}_{\mathbf{h},\mathbf{m}}} \bar{D}$, subject to the average power constraint $\bar{P} \leq P$. For a set of static parallel AWGN channels and Gaussian vector sources, the optimal linear encoding matrix transmits one measurement over each channel [6]. The optimal mapping between channels and measurements is as follows: We first reorder the measurement vector \mathbf{s} to obtain $\bar{\mathbf{s}} = [s_{(1)}, \dots, s_{(N)}]$, such that $\sigma_{m_{(1)}}^2 \leq \sigma_{m_{(2)}}^2 \leq \dots \leq \sigma_{m_{(N)}}^2$, and reorder the channel vector \mathbf{h} to obtain $\bar{\mathbf{h}} = [h_{(1)}, \dots, h_{(N)}]$, such that $|h_{(1)}| \leq |h_{(2)}| \leq \dots \leq |h_{(N)}|$. Then, the optimal linear encoding matrix $\mathbf{F}_{\mathbf{h},\mathbf{m}}^*$ is diagonal with entries $[f_{(1)}(h_{(1)}, m_{(1)}), \dots, f_{(N)}(h_{(N)}, m_{(N)})]$, and it maps the ordered measurements to ordered channel states. In order to find the diagonal entries of $\mathbf{F}_{\mathbf{h},\mathbf{m}}^*$, we can explicitly rewrite the convex optimization problem by using the optimal mappings derived in [6], as follows:

$$\begin{aligned} \bar{D}^* \triangleq \min_{f_{(t)}} \mathbb{E}_{M_{(t)}, H_{(t)}} & \left[\frac{1}{N} \sum_{t=1}^N \frac{\sigma_{m_{(t)}}^2}{|h_{(t)}|^2 f_{(t)}(h_{(t)}, m_{(t)})^2 \sigma_{m_{(t)}}^2 + 1} \right] \\ \text{s.t. } \mathbb{E}_{M_{(t)}, H_{(t)}} & \left[\frac{1}{N} \sum_{t=1}^N f_{(t)}(h_{(t)}, m_{(t)})^2 \sigma_{m_{(t)}}^2 \right] \leq P, \end{aligned} \quad (13)$$

where the expectation is taken over $M_{(t)}$ and $H_{(t)}$ for $t \in [1:N]$. The t -th smallest entry of the requested parameter vector $\mathbf{m} = [m_1, m_2, \dots, m_N]$, is denoted by the r.v. $M_{(t)} \in [1:N]$ with the order statistics $p_{M_{(t)}}(m)$. Without loss of generality, we assume that ordering the entries of \mathbf{m} in ascending order, i.e., $m_{(1)} \leq m_{(2)} \leq \dots \leq m_{(N)}$, implies ordering the entries of the measurement vector \mathbf{s} in ascending variances, i.e., $\sigma_{m_{(1)}}^2 \leq \sigma_{m_{(2)}}^2 \leq \dots \leq \sigma_{m_{(N)}}^2$. Similarly, the t -th smallest entry of the channel vector $\mathbf{h} = [h_1, h_2, \dots, h_N]$ is denoted by the r.v. $H_{(t)} \in \mathbb{R}$ with the order statistics $p_{H_{(t)}}(h)$.

The optimal linear encoding matrix $\mathbf{F}_{\mathbf{h},\mathbf{m}}^*$ with diagonal entries $f_{(t)}^*(h_{(t)}, m_{(t)})$ for $t \in [1:N]$, can be found from the Lagrange and the KKT conditions as follows:

$$f_{(t)}^*(h_{(t)}, m_{(t)}) = \sqrt{\left[\frac{\delta^*}{|h_{(t)}| \sigma_{m_{(t)}}} - \frac{1}{|h_{(t)}|^2 \sigma_{m_{(t)}}^2} \right]^+}, \quad (14)$$

where δ^* is the optimal Lagrange multiplier, such that $\bar{P} = P$ in (13).

Similarly, the optimal power allocation and the corresponding distortion can be found by using the *water-filling reflected on a reciprocal mirror* interpretation. The optimal Lagrange multiplier δ^* depends on $p_{M(t)}(m)$ and $p_{H(t)}(h)$, which can be found explicitly by using the order statistics. In the following lemma, we give the t -th order statistics $p_{M(t)}(m)$ and $p_{H(t)}(h)$, for $t \in [1:N]$.

Lemma 2. *Let $F_M(m)$ and $F_H(h)$ denote the cumulative distribution functions of $p_M(m)$ and $p_H(h)$, respectively. Given $F_M(m)$, $p_M(m)$, $F_H(h)$, $p_H(h)$ and N , the t -th order statistics $p_{M(t)}(m)$ and $p_{H(t)}(h)$, $t \in [1:N]$, are found as:*

$$p_{H(t)}(h) = t p_H(h) \binom{N}{t} (F_H(h))^{t-1} (1 - F_H(h))^{N-t}, \quad (15)$$

$$p_{M(t)}(m) = \sum_{b=t}^N \binom{N}{b} [F_M(m)^b (1 - F_M(m))^{N-b} - F_M(m-1)^b (1 - F_M(m-1))^{N-b}]. \quad (16)$$

Proof: The proof is trivial and achieved through the definition of the cumulative distribution functions of $H(t)$ and $M(t)$.

$$F_{H(t)}(h) = \Pr\{H(t) \leq h\} = \Pr\{\text{at least } t \text{ of } H\text{'s are } \leq h\}, \quad (17)$$

$$= \sum_{b=t}^N \frac{N!}{(N-b)!b!} F_H(h)^b (1 - F_H(h))^{N-b}. \quad (18)$$

where (17) implies a binomial distribution with at least t successes and can be formulated as (18). The t -th order statistics $p_{H(t)}(h)$ is found by taking the derivative of (18) with respect to h . The same proof holds for $M(t)$. ■

IV. LT STRATEGIES

In this section, we propose two LT strategies for general delay constraints $d \geq 1$. The block diagram of the proposed LT strategies is illustrated in Fig. 3. Both strategies are composed of two main blocks, namely, storage and transmission blocks. There are two buffers of size \bar{d} measurements, namely, the measurement buffer (MB) and the transmission buffer (TB). Here, we present these two schemes for an odd delay constraint, i.e., $d \in \{1, 3, 5, \dots\}$, but they can

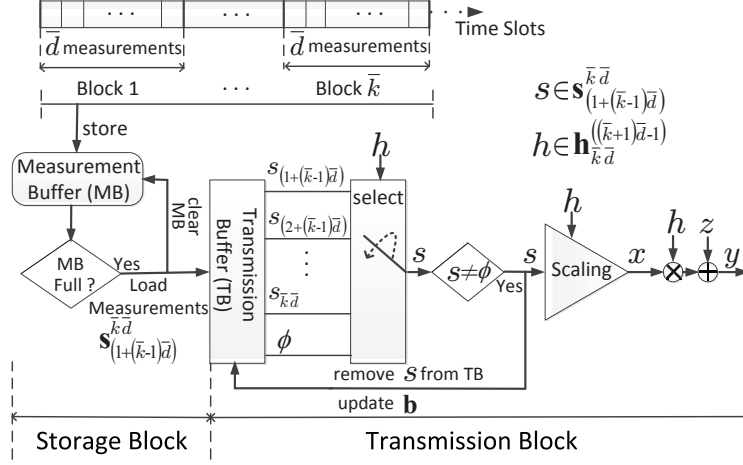


Fig. 3. The block diagram illustration of the proposed LT strategies.

be easily adapted to the case when d is even. In the storage block, given a delay constraint of $d = 2\bar{d} - 1$ for $\bar{d} \in [1:\infty]$, the sensor collects a block of \bar{d} consecutive measurements after \bar{d} consecutive TSs, and stores them in the MB. The consecutive blocks of \bar{d} measurements, taken over successive time intervals, are indexed by $\bar{k} = \{1, 2, \dots\}$. Then, the \bar{k} -th block consists of the measurements taken within TSs $[(1 + (\bar{k} - 1)\bar{d}) : \bar{k}\bar{d}]$, i.e., $s_{(1+(\bar{k}-1)\bar{d})}^{\bar{k}\bar{d}}$. When the MB gets full with the \bar{d} measurements of the \bar{k} -th block, the sensor removes $s_{(1+(\bar{k}-1)\bar{d})}^{\bar{k}\bar{d}}$ from the MB and loads them into the TB. Then, for the next consecutive \bar{d} TSs $[\bar{k}\bar{d} : ((\bar{k} + 1)\bar{d} - 1)]$, the sensor accesses the channel and transmits a linear function of the measurements in the TB, i.e., $s_{(1+(\bar{k}-1)\bar{d})}^{\bar{k}\bar{d}}$, over the channel states $\mathbf{h}_{\bar{k}\bar{d}}^{((\bar{k}+1)\bar{d}-1)}$ satisfying the delay constraint d . The specifics of these linear functions will be explained below.

Note that, while the sensor transmits the measurements in the TB, it starts refilling the MB with new measurements $s_{(\bar{k}\bar{d}+\bar{d})}^{(\bar{k}\bar{d}+\bar{d})}$. After \bar{d} channel accesses within TSs $[\bar{k}\bar{d} : ((\bar{k} + 1)\bar{d} - 1)]$, the MB gets full again and its new \bar{d} measurements are transferred to the TB for transmission over the next \bar{d} TSs.

The proposed transmission strategies consist of two sub-blocks, namely, the measurement selection and scaling sub-blocks. This division is motivated by the results of [6] presented in Section III-A, in which N ordered measurements are mapped one-to-one to N ordered channels, and each measurement is transmitted over its corresponding channel. Hence, we assume that, at each channel access, the sensor selects only one measurement and scales it to a channel input

value. However, in this case, we cannot directly use the optimal LT scheme in [6] and guarantee that the selected measurement and the channel state satisfy the optimal matching. This is because even though \bar{d} measurements are available in the TB in advance, the states of the next \bar{d} channels are not available to the transmitter as in the parallel channel model of [6]; and instead, they become available over time. The two proposed LT strategies differ in the way they choose the measurement to be transmitted at each TS.

A. Linear Transmission Scheme with Hard Matching (LTHM)

This transmission scheme has the following measurement selection criteria. Assume, without loss of generality, that parameters are ordered such that $\sigma_1^2 > \sigma_2^2 > \dots > \sigma_J^2$. We divide the channel magnitude space (\mathbb{R}^+) into J ordered channel intervals as, $\mathcal{H}_m = [H'_m, H'_{(m-1)})$, where $H'_m < H'_{(m-1)}$ for $m \in [1:J]$. The boundary values are chosen as $H'_0 = \infty$, $H'_J = 0$ and $H'_m = F_H^{-1}(1 - \sum_{j=1}^m p_M(j))$, for $m \in [1:(J-1)]$, where $F_H^{-1}(\cdot)$ denotes the inverse of the cumulative distribution function of the channel magnitude $|h|$, $F_H(|h|)$. Observe that according to this choice, the probability of the channel magnitude belonging to \mathcal{H}_m is $\Pr\{|h| \in \mathcal{H}_m\} = p_M(m)$.¹

The algorithmic description of LTHM is given in Algorithm 1. Let $\mathbf{b} = [b_1, b_2, \dots, b_J]$ be a J -length vector, where the m -th entry, $b_m \in [0:\bar{d}]$, denotes the number of measurements of parameter m in the TB, for $m \in [1:J]$. At each channel access, if $|h| \in \mathcal{H}_m$ and $b_m \neq 0$, then the sensor selects one measurement of the parameter type m from the TB and feeds it to the scaling sub-block. If there are multiple measurements of the same parameter type m in the TB, i.e., $b_m > 1$, then the sensor selects one of them randomly. The selected measurement is removed from the TB and \mathbf{b} is updated by reducing the m -th entry, b_m , by one. Thereby, each measurement is transmitted only once. On the other hand, if $|h| \in \mathcal{H}_m$ and $b_m = 0$, no measurement is transmitted in that TS. Hence, LTHM considers a hard matching condition for selecting measurements, in which each parameter has a corresponding interval of channel states,

¹If channel fading follows a discrete distribution, we define sets of channel states as opposed to intervals. With abuse of notation, we denote the m th set as \mathcal{H}_m , for $m \in [1:J]$. Suppose that the discrete channel states are ordered as $|\hat{h}_1| > |\hat{h}_2| > |\hat{h}_3| > \dots$. We allocate the discrete states into J sets such that the probability of channel state falling into set \mathcal{H}_m is $p_M(m)$. However, it may be possible that the channel states cannot be grouped to satisfy this equality exactly for all m . In that case we create virtual states to satisfy these equalities, as explained below.

Let j be the minimum index for which $\sum_{i=1}^j p_H(|h| = \hat{h}_i) > p_M(1)$. Define $p_M^1 = p_M(1) - \sum_{i=1}^{j-1} p_H(|h| = \hat{h}_i)$. We define a new virtual channel state \hat{h}_j^1 , whose gain is equivalent to \hat{h}_j . Whenever the real channel state is \hat{h}_j , we randomly assign the channel state to \hat{h}_j^1 with probability $p_M^1/p_M(j)$. We let $\mathcal{H}_1 = \{\hat{h}_1, \dots, \hat{h}_{j-1}, \hat{h}_j^1\}$. We repeat the same process for $p_M(2)$, starting with channel state \hat{h}_j whose probability is now $p_H(\hat{h}_j) - p_M^1$.

and only measurements of that parameter can be transmitted over a channel state from that interval. Note that, since the channel state is known at the receiver, it also knows which type of measurement is transmitted at each TS.

For the scaling sub-block we use the power allocation strategy derived in Section III. Thus, the selected measurement of the parameter type m is transmitted at the current channel state $|h| \in \mathcal{H}_m$, for $m \in [1:J]$, by allocating power $P(h, m)$, leading to distortion $D(h, m)$:

$$P(h, m) = \begin{cases} \left[\frac{\mu \sigma_m}{|h|} - \frac{1}{|h|^2} \right]^+, & \text{if hard matching holds,} \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

$$D(h, m) = \begin{cases} \frac{\sigma_m^2}{|h|^2 P(h, m) + 1}, & \text{if hard matching holds,} \\ \sigma_m^2, & \text{otherwise,} \end{cases} \quad (20)$$

where μ is chosen such that the average power constraint is satisfied.

After every transmission, the CC estimates the transmitted measurement s by using the channel output y . It is noteworthy that after \bar{d} channel accesses, we may have untransmitted measurements in the TB. TB is emptied anyway since these measurements have expired, and they are estimated with the maximum distortion σ_m^2 . As we show next, the average number of untransmitted measurements decreases with the increasing delay constraint d . However, for a finite delay constraint the untransmitted measurements dominate the distortion even for a high average transmission power constraint. In order to combat this drawback, we propose an alternative LT scheme.

B. Linear Transmission Scheme with Soft Matching (LTSM)

The algorithmic description of LTSM is given in Algorithm 1. The LTSM retains the hard matching condition of LTHM, i.e., at each channel access, if $|h| \in \mathcal{H}_m$ and $b_m \neq 0$ for $m \in [1:J]$, LTSM selects one measurement of the parameter type m from the TB. Hence, LTSM also gives the highest selection priority to the measurement of the parameter type that satisfies the hard matching condition with the channel state. However, if $|h| \in \mathcal{H}_m$ and $b_m = 0$, LTSM does not waste the channel state; and instead, selects one measurement based on the following measurement selection criteria:

Algorithm 1 LTHM and LTSM

```

1: Initialization:
   Load measurements of MB,  $\mathbf{s}_{(1+(\bar{k}-1)\bar{d})}^{\bar{k}\bar{d}}$ , into TB and update  $\mathbf{b}$ .
2: for  $i = \bar{k}\bar{d}$  to  $(\bar{k} + 1)\bar{d} - 1$  do                                ▷ TSs for  $\bar{d}$  channel accesses
3:   if  $|h_i| \in \mathcal{H}_m$  and  $b_m \neq 0$  then                                ▷ both for LTHM and LTSM
   Select one measurement of parameter  $m$  from TB.                        ▷ measurement selection
   Transmit the measurement over  $|h_i|$  with an allocated power of Eqn. (19).  ▷ scaling
    $b_m \leftarrow b_m - 1$                                                     ▷ update  $\mathbf{b}$ 
4:   else                                                                    ▷ only for LTSM
   Find  $\varsigma$  by solving  $\min_{b_\varsigma \neq 0} ||h_i| - h'_\varsigma|$ .
   Select one measurement of parameter  $\varsigma$  from TB.                        ▷ measurement selection
   Transmit the measurement over  $|h_i|$  with an allocated power of Eqn. (19).  ▷ scaling
    $b_m \leftarrow b_m - 1$                                                     ▷ update  $\mathbf{b}$ 
5:   end if
6: end for
    $\bar{k} \leftarrow \bar{k} + 1$  and go to Initialization

```

Assume that each interval \mathcal{H}_m is further divided into two equally probable intervals by the boundary value $h'_m = F_H^{-1} \left(\frac{F_H(H'_{(m-1)}) + F_H(H'_m)}{2} \right)$, for $\forall m \in [1:J]^2$. If $|h| \in \mathcal{H}_m$ and $b_m = 0$, then LTSM selects one measurement of parameter ς , which is the parameter that minimizes the following distance metric:

$$\min_{b_\varsigma \neq 0} ||h| - h'_\varsigma|. \quad (21)$$

When the hard matching condition is not satisfied, the LTSM considers a soft matching condition for selecting measurements; that is, among all parameter types of the measurements in the TB, it selects a measurement of the parameter whose corresponding interval of channel states has the value h'_ς closest to the channel state magnitude $|h|$. If two distinct ς values satisfy the solution of Eqn. (21), then LTSM chooses the smallest value of ς . LTSM allocates the power as in Eqn. (19), and transmits the selected measurement, leading to distortion in Eqn. (20). Note that the optimal Lagrange multiplier μ is chosen such that the average power constraint is satisfied. At the end of \bar{d} channel accesses, the sensor will have transmitted all the measurements in the TB, albeit some might have been allocated zero power as a result of the water-filling algorithm.

²If the channel follows a discrete fading distribution, we find h'_m by taking the mean value of all elements of channel set \mathcal{H}_m .

V. DISTORTION LOWER BOUNDS

We characterize two lower bounds on the MSE distortion, namely, the TLB and the LLB. While the TLB is the theoretical performance bound derived without any delay or complexity constraints on the transmission, the LLB is a performance lower bound only for LT strategies. We also prove that the proposed LT strategies meet the TLB under infinite delay and certain matching conditions between the channel states and parameter variances.

A. Theoretical Lower Bound (TLB)

Shannon's source-channel separation theorem states that the optimal end-to-end distortion is achieved by concatenating the optimal source and channel codes when there is no delay or complexity constraints, and the source and channel distributions are ergodic [17]. When we remove the delay and linear encoding constraints in our system model, then the sensor can transmit to the CC at the ergodic capacity, \bar{C}_e , of the underlying fading channel, while the minimum distortion, \bar{D}_e , is found by evaluating the distortion-rate function for a composite Gaussian source model at the ergodic capacity.

Since the channel state is known by both the transmitter and receiver, the ergodic capacity, in terms of the optimal power allocation scheme $P_e^*(h)$, is given by:

$$\bar{C}_e \triangleq \mathbb{E}_H \left[\frac{1}{2} \log (1 + |h|^2 P_e^*(h)) \right], \quad (22)$$

where $P_e^*(h)$ is found by the water-filling algorithm as $P_e^*(h) = [\alpha^* - 1/|h|^2]^+$, where α^* is chosen to satisfy $\bar{P}_e \triangleq \mathbb{E}_H [P_e^*(h)] = P$.

From Eqn. (6.1.21) of [14], the distortion-rate function of a composite Gaussian source with m components, $\mathcal{N}(0, \sigma_m^2)$, each of which is observed with probability $p_M(m)$ for $m \in [1:J]$, is defined as:

$$\bar{D}_e \triangleq \mathbb{E}_M [\sigma_m^2 2^{-2R_e^*(\sigma_m)}], \quad (23)$$

where the optimal rate allocated to source m , $R_e^*(\sigma_m)$, and the corresponding distortion, $D_e^*(\sigma_m)$, are given by:

$$R_e^*(\sigma_m) = \frac{1}{2} \left[\log \left(\frac{\sigma_m^2}{\beta^*} \right) \right]^+, \quad (24)$$

$$D_e^*(\sigma_m) = \min(\beta^*, \sigma_m^2), \quad (25)$$

where β^* is chosen such that $\bar{R}_e \triangleq \mathbb{E}_M[R_e^*(\sigma_m)] = \bar{C}_e$.

Hence, the optimal distortion is found as $\bar{D}_e = \mathbb{E}_M[D_e^*(\sigma_m)]$, which is the TLB on the achievable MSE distortion by any transmission strategy. Note that we have removed both the delay constraint and the linearity requirement on the encoder and decoder.

1) *Asymptotic Optimality of LT*: In general, the TLB cannot be achieved by LT strategies even if the delay constraint is removed. However, it can be shown that LTHM and LTSM meet this lower bound when the delay constraint is removed under certain matching conditions between the channel states and the parameter variances.

Assume that the channel follows a discrete fading distribution, where the channel state h can take one of the J values \hat{h}_m with probability $p_H(\hat{h}_m)$ for $m \in [1:J]$. The discrete values are ordered as $|\hat{h}_1| > |\hat{h}_2| > \dots > |\hat{h}_J|$. The next theorem states the necessary conditions in this discrete channel model under which LTHM and LTSM achieve the optimal distortion performance when the delay constraint is removed.

Theorem 1. *For the discrete AWGN fading channel model, if the parameter variances and the discrete channel states satisfy $\frac{\sigma_1}{|\hat{h}_1|} = \dots = \frac{\sigma_J}{|\hat{h}_J|}$, and $p_M(m) = p_H(\hat{h}_m)$, for $\forall m \in [1:J]$, then the TLB is achieved by LTHM and LTSM when the delay constraint is removed, i.e., $d \rightarrow \infty$.*

Proof: The proof can be found in Appendix A. ■

B. The Linear Transmission Lower Bound (LLB)

We next derive a lower bound on the achievable MSE distortion as a function of the delay and power constraints for any LT strategy. In order to derive this lower bound, we relax the assumption on the causal knowledge of the measurements and channel states, and instead assume that the sensor has the offline (non-causal) knowledge of a certain number of future measurements and channel states. Accordingly, we assume that at any TS the sensor non-causally knows the length- \bar{u} measurement vector, i.e., $\mathbf{s} = [s_1, \dots, s_{\bar{u}}]$, taken over the next \bar{u} TSs. Observe that, for a delay constraint d , each measurement of \mathbf{s} can only be transmitted over the following d channel states observed after it is taken, thus the transmission of the vector \mathbf{s} spans the following $\bar{c} = (d + \bar{u} - 1)$ channel states observed after the first measurement s_1 is taken. We further assume

that the sensor non-causally knows the length- \bar{c} channel vector $\mathbf{h} = [h_1, \dots, h_{\bar{c}}]$. Henceforth, the problem is reduced to optimally transmitting \bar{u} measurements over \bar{c} parallel channels, which is attained by using the optimal LT scheme presented in Section III-A. Accordingly, we first reorder \mathbf{s} to get $\bar{\mathbf{s}} = [s_{(1)}, \dots, s_{(\bar{u})}]$, where the variances of the ordered measurements satisfy $\sigma_{m_{(1)}}^2 \leq \sigma_{m_{(2)}}^2 \leq \dots \leq \sigma_{m_{(\bar{u})}}^2$, and reorder \mathbf{h} to get $\bar{\mathbf{h}} = [h_{(1)}, \dots, h_{(\bar{c})}]$, such that the ordered fading states satisfy $|h_{(1)}| \leq |h_{(2)}| \leq \dots \leq |h_{(\bar{c})}|$. Then, the $\bar{c} \times \bar{u}$ optimal linear encoding matrix $\mathbf{F}_{\mathbf{h}, \mathbf{m}}^*$ consists of a $\bar{u} \times \bar{u}$ size diagonal partition with entries $[f_{(1)}(h_{(1+\bar{e})}, m_{(1)}), \dots, f_{(\bar{u})}(h_{(\bar{u}+\bar{e})}, m_{(\bar{u})})]$, and a $\bar{e} \times \bar{u}$ size partition with zero entries, where $\bar{e} = \bar{c} - \bar{u}$, and it maps \bar{u} ordered measurements to the \bar{u} channels with the largest gains. The optimal entries of $\mathbf{F}_{\mathbf{h}, \mathbf{m}}^*$ are found as the solution of the following convex optimization problem with the optimal objective function $\bar{D}^*(d, \bar{u}, P)$:

$$\begin{aligned} \min_{f_{(t)}} \mathbb{E}_{M_{(t)}, H_{(t+\bar{e})}} & \left[\frac{1}{\bar{u}} \sum_{t=1}^{\bar{u}} \frac{\sigma_{m_{(t)}}^2}{|h_{(t+\bar{e})}|^2 f_{(t)}(h_{(t+\bar{e})}, m_{(t)})^2 \sigma_{m_{(t)}}^2 + 1} \right] \\ \text{s.t. } \bar{P} \triangleq \mathbb{E}_{M_{(t)}, H_{(t+\bar{e})}} & \left[\frac{1}{\bar{u}} \sum_{t=1}^{\bar{u}} f_{(t)}(h_{(t+\bar{e})}, m_{(t)})^2 \sigma_{m_{(t)}}^2 \right] \leq P, \end{aligned} \quad (26)$$

where the expectation is taken over $M_{(t)}$ and $H_{(t+\bar{e})}$ for $t \in [1:\bar{u}]$. The t -th and $(t + \bar{e})$ -th order statistics $p_{M_{(t)}}(m)$ and $p_{H_{(t+\bar{e})}}(h)$, are given by Lemma 2. The optimal linear encoding matrix with diagonal entries is found as:

$$f_{(t)}^*(h_{(t+\bar{e})}, m_{(t)}) = \sqrt{\left[\frac{\zeta^*}{|h_{(t+\bar{e})}| \sigma_{m_{(t)}}} - \frac{1}{|h_{(t+\bar{e})}|^2 \sigma_{m_{(t)}}^2} \right]^+}, \quad (27)$$

where ζ^* is the optimal Lagrange multiplier, such that $\bar{P} = P$ in (26).

Assuming non-causal knowledge of \bar{u} measurements and \bar{c} channel states under the delay constraint d and the average power constraint P , we obtain the optimal distortion $\bar{D}^*(d, \bar{u}, P)$ for any LT strategy. Then, the LLB is derived by finding the \bar{u} value, which maximizes $\bar{D}^*(d, \bar{u}, P)$:

$$\bar{D}_l(d, P) \triangleq \max_{\bar{u}} \bar{D}^*(d, \bar{u}, P). \quad (28)$$

Note that we have relaxed the constraint for the causal knowledge of measurements and channel states both at the encoder and decoder. The numerical comparisons of the LLB with the proposed schemes will be presented in Section VII.

VI. NO CSI AT THE ENCODER

In this section, we assume that the CSI is known only at the decoder. We derive the optimal LT strategy under a strict delay constraint ($d = 1$), as well as the TLB on the achievable MSE distortion. Additionally, for the multiple measurements-parallel channels scenario studied in Section III-A, we show that if the CSI is available only at the receiver, any LT scheme that is limited to a one-to-one linear mapping from the measurements to the channel input is suboptimal in general. The optimal LT strategy is elusive and it will be a non-trivial function of the source variances and the channel distribution.

A. Strict Delay Constraint

Under a strict delay constraint, the most recent measurement is transmitted at each TS. By applying Lemma 1 to this scenario, we can similarly show that there is no loss of optimality by considering time-invariant encoding functions, i.e., $f_i(m) = f(m)$, $\forall i$. Hence, the encoding function $f(m)$ is a scalar and time-invariant. The decoding function $g(h, m)$ that minimizes the MSE is the linear MMSE estimator [15], and is also a scalar and time-invariant. Then, the MSE distortion, $\bar{D} = \mathbb{E}_{M,H,S,Z}[|S - \hat{S}|^2]$, and the average power, $\bar{P} = \mathbb{E}_{M,S}[|X|^2]$, can be written explicitly as:

$$\bar{D} = \sum_{m=1}^J p_M(m) \int_{\mathbb{R}} \frac{\sigma_m^2}{|h|^2 f(m)^2 \sigma_m^2 + 1} p_H(h) dh, \quad (29)$$

$$\bar{P} = \sum_{m=1}^J p_M(m) f(m)^2 \sigma_m^2. \quad (30)$$

where $P(m) \triangleq f(m)^2 \sigma_m^2$. The optimal linear encoding function, $f^*(m)$, is found as the solution to the convex optimization problem $\bar{D}^* \triangleq \min_f \bar{D}$, subject to the average power constraint $\bar{P} \leq P$. From the KKT conditions [16], we have:

$$f^*(m) = \sqrt{\frac{\left[\Psi^{-1}\left(\frac{\lambda^*}{\sigma_m^2}\right) \right]^+}{\sigma_m^2}}, \quad (31)$$

where $\Psi^{-1} : \mathbb{R} \rightarrow \mathbb{R}$ is the inverse of the function $\Psi : \mathbb{R} \rightarrow \mathbb{R}$, that is defined as, $\Psi(P(m)) \triangleq \int_{\mathbb{R}} \frac{|h|^2}{(|h|^2 P(m) + 1)^2} p_H(h) dh$. The optimal Lagrange multiplier λ^* is chosen such that $\bar{P} = P$ in (30).

B. Multiple Measurements and Parallel Channels

Next we consider the multiple measurements-parallel channels scenario studied in Section III-A, under the strict delay constraint and the assumption that the CSI is known only at the decoder, and $J > 1$. In such a scenario, the optimal LT scheme of [6], in which the ordered measurements are mapped one-to-one to ordered channel states, cannot be used directly. This is because, even though the encoder knows the N measurements, it does not know any of the channel states, and hence; cannot order them. For the special case where N measurements are observed from a single Gaussian source ($J = 1$), in [7] the authors show that the optimal performance is achieved by transmitting one measurement over each channel. When $J = 1$, since N measurements all have the same variance, all orderings are equivalent, and the optimal LT performance is achieved by an LT scheme that uses only a one-to-one mapping between measurements and channels. However, this is not the case in general when $J > 1$. Since N measurements follow a composite Gaussian source model, the encoder can have measurements with different variances; and hence, we can exploit the diversity of the fading channel by transmitting a single measurement over multiple channels, instead of transmitting each measurement only once. Depending on the source variances, the former may surpass the best LT performance achieved by using only a one-to-one linear mapping. This is shown in the following lemma by considering a particular example.

Lemma 3. *Consider the LT of N measurements of a composite Gaussian source with $J > 1$ components over N parallel AWGN fading channels. If the CSI is known only by the decoder, then the LT scheme that uses a one-to-one linear mapping between measurements and channels is suboptimal in general.*

Proof: The proof can be found in Appendix B. ■

C. The Theoretical Lower Bound (TLB)

Similarly to Section V-A, we derive the TLB on the achievable MSE distortion by using Shannon's source-channel separation theorem. If the CSI is available only at the decoder and the average power constraint is P , then the ergodic capacity is given by:

$$\bar{C}_e \triangleq \mathbb{E}_H \left[\frac{1}{2} \log (1 + |h|^2 P) \right]. \quad (32)$$

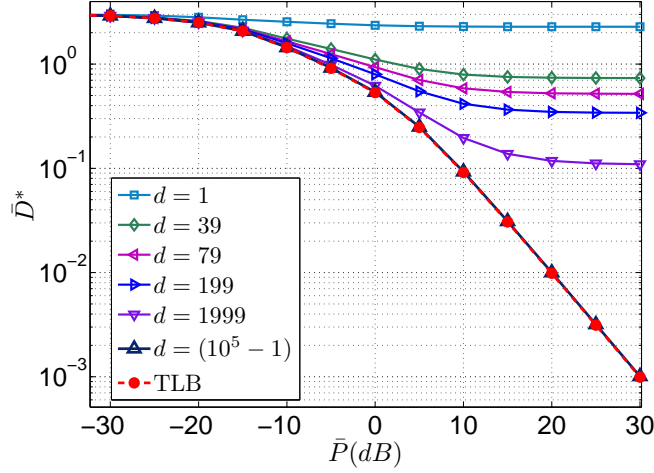


Fig. 4. Achievable MSE distortion with LTHM with respect to average power for different delay constraints in the discrete fading channel model.

The distortion-rate function of a composite Gaussian source is defined as in Eqn. (23) of Section V-A, which leads to the optimal rate allocated to source m , $R_e^*(\sigma_m)$, as in Eqn. (24) and the corresponding distortion, $D_e^*(\sigma_m)$, as in Eqn. (25), respectively. The Lagrangian multiplier β^* for this case is chosen such that $E_M[R_e^*(\sigma_m)]$ is equal to the ergodic capacity \bar{C}_e in (32). Then the TLB on the achievable MSE distortion by any strategy when the encoder does not have the CSI is given by $\bar{D}_e = E_M[D_e^*(\sigma_m)]$.

VII. NUMERICAL RESULTS AND OBSERVATIONS

Here we provide numerical results to compare the performances of LTHM and LTSM with the lower bounds, and to analyze the impact of the delay and power constraints on the performance. In our simulations, we consider $J = 4$ Gaussian parameters with variances $\{10, 5, 1, 0.5\}$, which are requested with probabilities $\{0.1, 0.3, 0.4, 0.2\}$, respectively. For a continuous fading channel, we consider Rayleigh distribution with a scale parameter $\omega = 3$, and for a discrete fading channel, we consider four states $\{\sqrt{10}, \sqrt{5}, 1, \sqrt{0.5}\}$, which are observed with probabilities $\{0.1, 0.3, 0.4, 0.2\}$, respectively.

We illustrate the achievable MSE distortion versus average power under various delay constraints with LTHM in the discrete channel setting in Fig. 4. We observe that the MSE distortion diminishes as the delay constraint is relaxed. This is because a relaxed delay constraint provides a

larger number of measurements in the TB; and hence, more flexibility for the sensor in selecting the appropriate measurement for each TS. We note that this statement does not hold when $J = 1$, in which case increasing the block length does not provide any improvement [6]. As it can be seen in Fig. 4, the MSE distortion converges to a fixed value as the average power value increases. This is due to the additional distortion brought in by the untransmitted measurements in the TB. The average number of untransmitted measurements and their effect on the MSE distortion decreases as the delay constraint is relaxed, since having a larger number of measurements in the TB increases the probability of finding a measurement that satisfies the hard matching condition. In particular, when the delay constraint is removed, as seen in Fig. 4, LTHM achieves the TLB, and becomes the optimal LT scheme, since the source-channel matching conditions in Theorem 1 are satisfied for the setup considered here.

In Fig. 5, we illustrate the achievable MSE distortion with LTSM with respect to average power under various delay constraints in the continuous channel model. Similarly to LTHM, the MSE distortion diminishes as the delay constraint increases. On the other hand, as opposed to LTHM, the MSE distortion achieved by LTSM decreases monotonically with the average power as illustrated in Fig. 5. This is because the performance of LTSM does not suffer from a fixed distortion component due to the untransmitted measurements. In addition, LTSM also approaches the TLB as the delay constraint is relaxed. Although we do not expect the LTSM to meet the TLB in this setting since the matching conditions of Theorem 1 do not hold, we observe in Fig. 5 that it is very close to the TLB.

Next, we compare the performances of LTHM and LTSM with each other and with the TLB and the LLB. Fig. 6 shows the achievable MSE distortion of LTHM, LTSM, the LLB and the TLB with respect to delay constraint in the continuous fading channel model for an average power constraint $\bar{P} = 10$ dB. As seen in the figure, the performance of the TLB is constant since it is derived by completely removing the delay and complexity constraints. On the other hand, the LLB decays slowly as the delay constraint increases. As expected, the MSE distortion of LTHM and LTSM decrease as the delay constraint increases. We can see that LTSM meets the LLB under the strict delay constraint. As expected, LTSM always outperforms LTHM, while the gap between the two schemes decreases with the increasing delay constraint. The gap between the TLB and two schemes also decreases with the increasing delay constraint even though we do not expect either of the schemes converge to the TLB in this setting since the matching

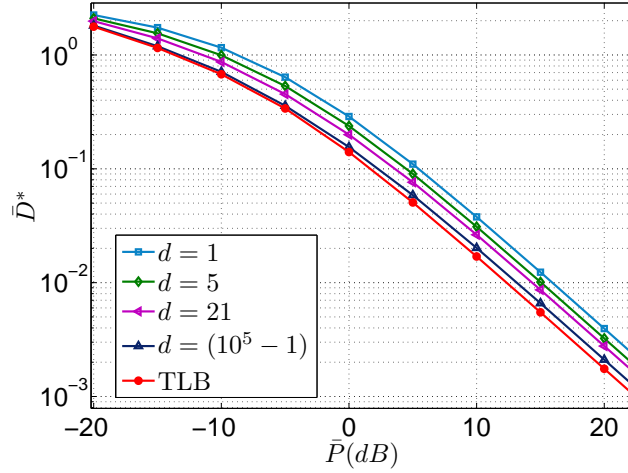


Fig. 5. Achievable MSE distortion with LTSM with respect to average power for various delay constraints in the continuous fading channel model.

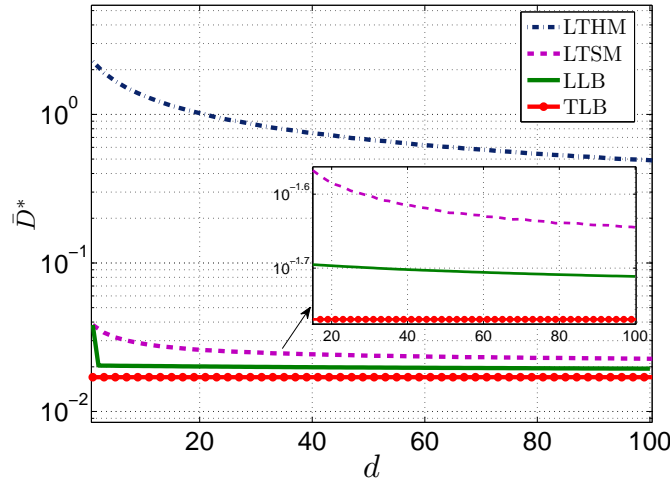


Fig. 6. MSE distortion versus delay constraint, d , in the continuous fading channel model for an average power constraint $\bar{P} = 10$ dB.

conditions of Theorem 1 do not hold.

Finally, in Fig. 7, we illustrate the achievable MSE distortion of LT and the TLB with respect to average power in the discrete channel model for the scenarios in which the CSI is known only by the decoder, and by both the encoder and decoder. The MSE distortion of LT under strict delay constraint of $d = 1$ for both scenarios diminishes as the average power increases. However, there is a constant gap between the optimal performances achieved with and without

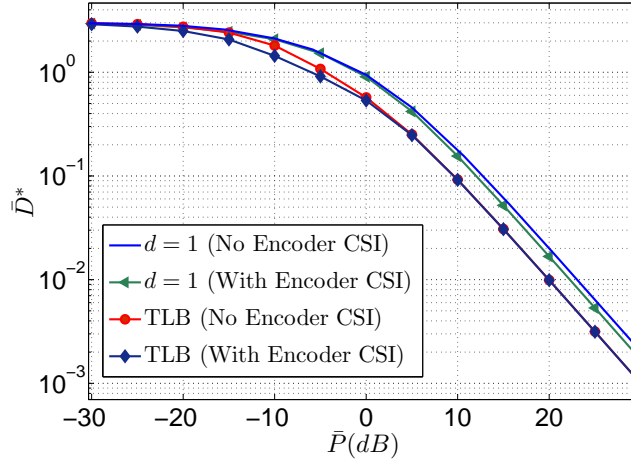


Fig. 7. The achievable MSE distortion of LT and the TLB with respect to average power in the discrete fading channel model with and without encoder CSI.

encoder CSI at higher \bar{P} values. On the other hand, the TLB for both scenarios meet as the average power increases since the gain from the optimal power allocation over different channel states disappears in the high power regime.

VIII. CONCLUSIONS

We have studied the delay-constrained LT of composite Gaussian measurements from a sensor to a CC over an AWGN fading channel. We have considered a wireless sensor that can collect measurements from J distinct Gaussian parameters. The CC asks for a measurement of a particular parameter from the sensor with a certain probability at each TS. In this framework, we have presented the optimal LT strategy under a strict delay constraint, and have given a graphical interpretation for the optimal power allocation scheme and the corresponding distortion value. Then, we have proposed two LT strategies, called LTHM and LTSM, under general delay constraints, and have provided numerical results to investigate the impact of the delay and average power constraints on the performance. We have seen that, if the number of parameters, J , is more than one, the MSE distortion decreases as the delay constraint is relaxed. We have also derived lower bounds on the achievable MSE distortion for generic and LT strategies. While LTSM outperforms LTHM at all delay constraints, we have shown analytically that both strategies meet the lower bound when the delay constraint is removed, under certain matching conditions between the parameter and the channel statistics.

We have also studied the scenario in which the CSI is known only by the decoder. We have presented the optimal LT strategy under a strict delay constraint. We have derived a TLB on the achievable MSE distortion by relaxing the delay constraint and the linearity requirement. We have also considered the multiple measurements-parallel channels scenario under a strict delay constraint, and have shown that the optimal LT performance cannot be achieved by using only a one-to-one linear mapping between measurements and channels, as opposed to the results derived in [6] and [7]. The design of the optimal LT strategy for the multiple measurements-parallel channels scenario for arbitrary delay constraints is elusive, and is left as future work.

APPENDIX A

PROOF OF THEOREM 1

Given a delay constraint $d = 2\bar{d} - 1$, let the r.v. \bar{Z}_m , $m \in [1:J]$, denote the total number of measurements of parameter m among \bar{d} measurements loaded into the TB. \bar{Z}_m follows a Binomial distribution with parameters \bar{d} and $p_M(m)$. Hence, the probability of having \bar{k} measurements of parameter m in the TB is given by $p_{\bar{Z}_m}(\bar{k}) = \Pr\{\bar{Z}_m = \bar{k}\} = \binom{\bar{d}}{\bar{k}} p_M(m)^{\bar{k}} (1 - p_M(m))^{\bar{d}-\bar{k}}$. Similarly, considering the discrete fading model presented in Section V-A1, let the r.v. \hat{Z}_m , $m \in [1:J]$, denote the total number of channels with state \hat{h}_m , after \bar{d} channel accesses. \hat{Z}_m also follows a Binomial distribution with parameters \bar{d} and $p_H(\hat{h}_m)$. Hence, the probability of observing \hat{k} channels with state \hat{h}_m is given by $p_{\hat{Z}_m}(\hat{k}) = \Pr\{\hat{Z}_m = \hat{k}\} = \binom{\bar{d}}{\hat{k}} p_H(\hat{h}_m)^{\hat{k}} (1 - p_H(\hat{h}_m))^{\bar{d}-\hat{k}}$.

Observe that after \bar{d} channel accesses, the number of transmitted measurements selected from the TB with LTHM is given by $\min\{\bar{Z}_m, \hat{Z}_m\}$. On the other hand, the number of untransmitted measurements remained in the TB is given by $[\bar{Z}_m - \hat{Z}_m]^+$. Then, the average power, \bar{P}_∞ , and the achievable MSE distortion, \bar{D}_∞ , of LTHM when $\bar{d} \rightarrow \infty$ are given by:

$$\bar{P}_\infty \triangleq \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} \sum_{m=1}^J \mathbb{E}_{\bar{Z}_m, \hat{Z}_m} \left[\min\{\bar{Z}_m, \hat{Z}_m\} \right] P(\hat{h}_m, m), \quad (33)$$

$$\bar{D}_\infty \triangleq \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} \sum_{m=1}^J \left\{ \mathbb{E}_{\bar{Z}_m, \hat{Z}_m} \left[[\bar{Z}_m - \hat{Z}_m]^+ \right] \sigma_m^2 + \mathbb{E}_{\bar{Z}_m, \hat{Z}_m} \left[\min\{\bar{Z}_m, \hat{Z}_m\} \right] D(\hat{h}_m, m) \right\}, \quad (34)$$

where the allocated power $P(\hat{h}_m, m)$ and the distortion $D(\hat{h}_m, m)$ are chosen as in Eqn. (19) and Eqn. (20), respectively.

In the rest of the proof, we use $p(m)$ to refer to the condition of Theorem 1, i.e., $p_M(m) = p_H(\hat{h}_m) = p(m)$, $\forall m$. Under this condition, the expected value and variance of \bar{Z}_m and \hat{Z}_m can be found as, $E[\bar{Z}_m] = E[\hat{Z}_m] = \bar{d} \cdot p(m)$ and $\text{Var}[\bar{Z}_m] = \text{Var}[\hat{Z}_m] = \sigma_{Z_m}^2 = \bar{d} \cdot p(m) \cdot (1 - p(m))$, respectively. Let $\epsilon > 0$ be any positive number. Then, the Chebyshev's inequality leads to the following inequalities, $\Pr\{|\bar{Z}_m - \bar{d} \cdot p(m)| \geq \epsilon \cdot \sigma_{Z_m}\} \leq \frac{1}{\epsilon^2}$ and $\Pr\{|\hat{Z}_m - \bar{d} \cdot p(m)| \geq \epsilon \cdot \sigma_{Z_m}\} \leq \frac{1}{\epsilon^2}$. We define the interval \mathcal{I} on the real line as, $\mathcal{I} = [\bar{d} \cdot p(m) - \epsilon \cdot \sigma_{Z_m}, \bar{d} \cdot p(m) + \epsilon \cdot \sigma_{Z_m}]$.

Next, we compute (33) and (34) by finding upper and lower bounds on the expectation terms under the matching condition. Observe that,

$$\lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} E_{\bar{Z}_m, \hat{Z}_m} \left[\min \left\{ \bar{Z}_m, \hat{Z}_m \right\} \right] \leq \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} E_{\bar{Z}_m, \hat{Z}_m} [\bar{Z}_m] = p(m). \quad (35)$$

We can also lower bound this term as,

$$\lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} E_{\bar{Z}_m, \hat{Z}_m} \left[\min \left\{ \bar{Z}_m, \hat{Z}_m \right\} \right], \quad (36)$$

$$\geq \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} E_{\bar{Z}_m, \hat{Z}_m} \left[\min \left\{ \bar{Z}_m, \hat{Z}_m \right\} \middle| \substack{\bar{Z}_m \in \mathcal{I}, \\ \hat{Z}_m \in \mathcal{I}} \right] \Pr\{\bar{Z}_m \in \mathcal{I}, \hat{Z}_m \in \mathcal{I}\}, \quad (37)$$

$$\geq \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} \left(\bar{d} p(m) - \epsilon \sigma_{Z_m} \right) \left(1 - \frac{1}{\epsilon^2} \right)^2, \quad (38)$$

$$= \lim_{\bar{d} \rightarrow \infty} \left(p(m) - \frac{\sqrt{p(m)(1-p(m))}}{\bar{d}^{\frac{1}{6}}} \right) \left(1 - \frac{1}{\bar{d}^{\frac{2}{3}}} \right)^2 = p(m), \quad (39)$$

where (37) follows from the law of total expectation; (38) follows from the definition of \mathcal{I} , and the Chebyshev's inequality; and (39) is obtained by setting $\epsilon = \bar{d}^{\frac{1}{3}}$. Since the upper and lower bounds in (35) and (39) are equal, we have shown that (35) converges to $p(m)$ as $\bar{d} \rightarrow \infty$.

Similarly,

$$\lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} \mathbb{E}_{\bar{Z}_m, \hat{Z}_m} \left[[\bar{Z}_m - \hat{Z}_m]^+ \right], \quad (40)$$

$$= \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} \left\{ \mathbb{E}_{\bar{Z}_m, \hat{Z}_m} \left[[\bar{Z}_m - \hat{Z}_m]^+ \middle| \substack{\bar{Z}_m \in \mathcal{I}, \\ \hat{Z}_m \in \mathcal{I}} \right] \Pr\{\bar{Z}_m \in \mathcal{I}, \hat{Z}_m \in \mathcal{I}\} \right. \\ \left. + \mathbb{E}_{\bar{Z}_m, \hat{Z}_m} \left[[\bar{Z}_m - \hat{Z}_m]^+ \middle| \substack{\bar{Z}_m \notin \mathcal{I} \\ \text{or} \\ \hat{Z}_m \notin \mathcal{I}} \right] \Pr\{\bar{Z}_m \notin \mathcal{I} \text{ or } \hat{Z}_m \notin \mathcal{I}\} \right\}, \quad (41)$$

$$\leq \lim_{\bar{d} \rightarrow \infty} \frac{1}{\bar{d}} \left\{ 2\epsilon \sigma_{Z_m} + \left(\frac{2}{\epsilon^2} + \frac{1}{\epsilon^4} \right) \bar{d} \right\}, \quad (42)$$

$$= \lim_{\bar{d} \rightarrow \infty} \left\{ \left(\frac{2\sqrt{p(m)(1-p(m))}}{\bar{d}^{\frac{1}{6}}} \right) + \left(\frac{2}{\bar{d}^{\frac{2}{3}}} + \frac{1}{\bar{d}^{\frac{4}{3}}} \right) \right\} = 0, \quad (43)$$

where (41) follows from the law of total expectation; (42) follows from the definition of \mathcal{I} , and the Chebyshev's inequality; and (43) is obtained by setting $\epsilon = \bar{d}^{-\frac{1}{3}}$. This proves that (40) indeed converges to zero as $\bar{d} \rightarrow \infty$. This also implies that as $\bar{d} \rightarrow \infty$, all selected measurements by the LTSM strategy satisfy the hard matching condition. Hence, LTSM and LTHM are equivalent in the asymptotic of $\bar{d} \rightarrow \infty$ under the matching condition of Theorem 1.

Finally, we can rewrite \bar{P}_∞ and \bar{D}_∞ for both LTHM and LTSM as:

$$\bar{P}_\infty = \sum_{m=1}^J \left[\mu^* q - \frac{1}{|\hat{h}_m|^2} \right]^+ p(m), \quad (44)$$

$$\bar{D}_\infty = \sum_{m=1}^J \left[\frac{\sigma_m^2}{|\hat{h}_m|^2 \left[\mu^* q - \frac{1}{|\hat{h}_m|^2} \right]^+ + 1} \right] p(m), \quad (45)$$

where we use $q \triangleq \frac{\sigma_m}{|\hat{h}_m|}$, $\forall m$, from Theorem 1, and μ^* is chosen to satisfy $\bar{P}_\infty = P$.

Next, we show that $(\bar{P}_\infty, \bar{D}_\infty)$ pair above, obtained under the conditions of Theorem 1, achieve the TLB pair (\bar{P}_e, \bar{D}_e) , derived in Section V-A. First, under the matching condition, observe that $\mu^* q = \alpha^*$, and thus, $\bar{P}_\infty = \bar{P}_e = P$. Moreover, under the matching condition, $\bar{R}_e = \bar{C}_e$ in TLB implies $\alpha^* = \frac{q^2}{\beta^*}$. Combining the two equalities, we obtain $\mu^* = \frac{q}{\beta^*}$. Substituting this into Eqn. (23) together with the matching condition, we can show that $\bar{D}_e = \sum_{m=1}^J \min \left(\frac{q}{\mu^*}, \sigma_m^2 \right) p(m) = \bar{D}_\infty$, which concludes the proof of Theorem 1.

APPENDIX B

PROOF OF LEMMA 3

In order to prove Lemma 3, we construct a counter-example. We argue that the achievable MSE distortion of a particular LT scheme that is not constrained to use only a one-to-one mapping between measurements and channels can be smaller than the minimum achievable MSE distortion of all possible LT schemes that use only a one-to-one mapping, i.e., a diagonal encoding matrix. Suppose we have $J = 2$ zero-mean Gaussian parameters with variances σ_1^2 and σ_2^2 , which are requested with probabilities $p_M(1) = p_1$ and $p_M(2) = p_2 = (1 - p_1)$, respectively, and assume an extreme case, where $\sigma_1^2 > 0$ and $\sigma_2^2 = 0$. Suppose we have a discrete fading channel with two states, which are observed with probabilities $p_{H_1}(\hat{h}_1) = p_1$ and $p_{H_2}(\hat{h}_2) = p_2$, respectively, and assume that the channel states are $\hat{h}_1 > 0$ and $\hat{h}_2 = 0$. We aim at linearly transmitting $N = 2$ measurements of parameters $m_1 \in [1:2]$ and $m_2 \in [1:2]$, over $N = 2$ channel states $h_1 \in \{\hat{h}_1, \hat{h}_2\}$ and $h_2 \in \{\hat{h}_1, \hat{h}_2\}$.

We first characterize the minimum achievable MSE distortion, \bar{D}_1 , for all possible LT schemes with a diagonal encoding matrix. According to Eqn. (10), the encoding function needs to satisfy the average power constraint P , i.e., $\frac{1}{2} [P_{11}p_1^2 + P_{12}p_1p_2 + P_{21}p_1p_2 + P_{22}p_2^2] = P$, where $P_{m_1m_2}$ is the allocated power for the pair of measurements of parameters m_1 and m_2 , respectively. We have $P_{22} = 0$, since $\sigma_2^2 = 0$. Then, by using Eqn. (12), the MSE distortion \bar{D}_1 can be written explicitly as follows:

$$\begin{aligned} \bar{D}_1 = \frac{1}{2} \left\{ p_1^2 \left(\mathbb{E}_{H_1} \left[\frac{\sigma_1^2}{|h_1|^2 \frac{P_{11}}{2} + 1} \right] + \mathbb{E}_{H_2} \left[\frac{\sigma_1^2}{|h_2|^2 \frac{P_{11}}{2} + 1} \right] \right) \right. \\ \left. + p_1p_2 \left(\mathbb{E}_{H_1} \left[\frac{\sigma_1^2}{|h_1|^2 P_{12} + 1} \right] + \mathbb{E}_{H_2} \left[\frac{\sigma_1^2}{|h_2|^2 P_{21} + 1} \right] \right) \right\}, \end{aligned} \quad (46)$$

$$= p_1^2 \left(p_1 \frac{\sigma_1^2}{|\hat{h}_1|^2 \frac{P_{11}}{2} + 1} + p_2 \sigma_1^2 \right) + \frac{p_1p_2}{2} \left(p_1 \frac{\sigma_1^2}{|\hat{h}_1|^2 P_{12} + 1} + p_1 \frac{\sigma_1^2}{|\hat{h}_1|^2 P_{21} + 1} + 2p_2 \sigma_1^2 \right), \quad (47)$$

where the minimum distortion is achieved by dividing the power, i.e., P_{11} , equally between measurements if two measurements are observed from parameter 1, i.e., $m_1 = m_2 = 1$. If one measurement is requested from each parameter, i.e., $(m_1 = 1, m_2 = 2)$ or $(m_1 = 2, m_2 = 1)$, then the minimum distortion is achieved by allocating the entire power, i.e., P_{12} or P_{21} , to the measurement of parameter 1, since $\sigma_2^2 = 0$.

Assuming the average power constraint P is satisfied as in the above scheme, we next consider

a particular LT scheme. This scheme uses a diagonal encoding matrix if both measurements are observed from the same parameter; otherwise, it uses a non-diagonal matrix, where the measurement of parameter 1 is transmitted over two channels. Then, from Eqn. (12), the MSE distortion \bar{D}_2 can be written as follows:

$$\bar{D}_2 = \frac{1}{2} \left\{ p_1^2 \left(E_{H_1} \left[\frac{\sigma_1^2}{|h_1|^2 \frac{P_{11}}{2} + 1} \right] + E_{H_2} \left[\frac{\sigma_1^2}{|h_2|^2 \frac{P_{11}}{2} + 1} \right] \right) + p_1 p_2 \left(E_{H_1, H_2} \left[\frac{\sigma_1^2}{(|h_1|^2 + |h_2|^2) \frac{P_{12}}{2} + 1} \right] + E_{H_1, H_2} \left[\frac{\sigma_1^2}{(|h_1|^2 + |h_2|^2) \frac{P_{21}}{2} + 1} \right] \right) \right\}, \quad (48)$$

$$= p_1^2 \left(p_1 \frac{\sigma_1^2}{|h_1|^2 \frac{P_{11}}{2} + 1} + p_2 \sigma_1^2 \right) + \frac{p_1 p_2}{2} \left(2 p_2^2 \sigma_1^2 + p_1^2 \frac{\sigma_1^2}{|h_1|^2 P_{12} + 1} + p_1^2 \frac{\sigma_1^2}{|h_1|^2 P_{21} + 1} + 2 p_1 p_2 \frac{\sigma_1^2}{|h_1|^2 \frac{P_{12}}{2} + 1} + 2 p_1 p_2 \frac{\sigma_1^2}{|h_1|^2 \frac{P_{21}}{2} + 1} \right), \quad (49)$$

where the minimum distortion can be achieved by dividing the power, i.e., P_{11} , equally between measurements if two measurements are observed from parameter 1, i.e., $m_1 = m_2 = 1$, similarly to the above scheme. If one measurement is requested from each parameter, i.e., $(m_1 = 1, m_2 = 2)$ or $(m_1 = 2, m_2 = 1)$, then this particular scheme divides the power, i.e., P_{12} or P_{21} , equally between two channels h_1 and h_2 for the transmission of the measurement of parameter 1, as seen in the term multiplied by $p_1 p_2$ in (48). If two measurements are observed from parameter 2, i.e., $m_1 = m_2 = 2$, then we do not allocate power, i.e., $P_{22} = 0$, since $\sigma_2^2 = 0$.

We can easily see that $\bar{D}_2 < \bar{D}_1$ for all P_{11} , P_{12} and P_{21} . This implies that the minimum achievable MSE distortion of LT schemes constrained to one-to-one mapping can be improved by utilizing non-diagonal encoding matrices, which concludes the proof of Lemma 3.

REFERENCES

- [1] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Mar.-Apr. 2009.
- [2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, p. 393422, Mar. 2002.
- [3] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. New York, NY: CRC Press, 2000.
- [4] M. Hefeeda and M. Bagheri, "Wireless sensor networks for early detection of forest fires," in *Proc. IEEE Mobile Adhoc and Sensor Systems*, Pisa, Italy, Oct. 2007.
- [5] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. on Inf. Theory.*, vol. 11, no. 4, pp. 558–567, Oct. 1965.

- [6] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. on Commun.*, vol. 24, no. 12, pp. 1283–1290, Dec. 1976.
- [7] J. J. Xiao, Z. Q. Luo, and N. Jindal, "Linear joint source-channel coding for Gaussian sources through fading channels," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, CA, Dec. 2006.
- [8] J. Geng, H. Li, and L. Lai, "Smart grid system state measurement estimation over wireless channels," in *Proc. 46th Inform. Sci. Sys. Conf.*, Princeton, NJ, Mar. 2012.
- [9] I. E. Aguerri and D. Gündüz, "Linear transmission of correlated Gaussian sources over MIMO channels," in *Proc. IEEE Int. Symp. on Wireless Commun. Systems (ISWCS)*, Ilmenau, Germany, Aug. 2013.
- [10] A. Kashyap, T. Basar, and R. Srikant, "Minimum distortion transmission of Gaussian sources over fading channels," in *Proc. IEEE Conf. Decision and Control*, Maui, Hawaii, Aug. 2013.
- [11] I. Esnaola, A. M. Tulino, and J. Garcia-Frias, "Linear analog coding of correlated multivariate Gaussian sources," *IEEE Trans. on Commun.*, vol. 61, no. 8, pp. 3438–3447, Aug. 2013.
- [12] O. Tan, D. Gündüz, and J. Gómez-Vilardebò, "Delay constrained linear transmission of smart grid state measurements," in *Proc. 8th IEEE Sensor Array Multich. Sig. Process. Wkshp.*, Coruña, Spain, Jun. 2014.
- [13] O. Tan, D. Gündüz, and J. Gómez-Vilardebò, "Delay constrained linear transmission of a mixture of Gaussian measurements over a fading channel," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, UK, Jun. 2015.
- [14] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice Hall, 1971.
- [15] A. E. Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [17] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.